REPORT DOCUMENTATION PAGE	Form Approved OMB NO. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggesstions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any oenalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
06-10-2014	Final Report	1-Apr-2013 - 31-May-2014

4. TITLE AND SUBTITLE	
Final Report: General Multidecision Theory: Hypothesis Testing and Changepoint Detection with Applications to Homeland Security	5a. CONTRACT NUMBER W911NF-13-1-0073
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 611102
6. AUTHORS Alexander Tartakovsky	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES	8. PERFORMING ORGANIZATION REPORT NUMBER
University of Southern California 3720 S. Flower Street Los Angeles, CA 90089 -0701	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	10. SPONSOR/MONITOR'S ACRONYM(S) ARO 11. SPONSOR/MONITOR'S REPORT NUMBER(S) 63223-MA.6

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

The objective was to develop general theories of sequential hypothesis testing and quickest change detection for complex multi-population stochastic models, as well as to apply these theories to automatic threat detection and classification with low false alarm and miss-classification rates. More specifically, we addressed complex stochastic models that include: multi-population/multi-channel models; multi-hypothesis scenarios; general stochastic models with non-stationary and dependent observations; prior uncertainty. Within a limited duration of the project (about 8 months) a background for a general theory of testing multiple composite hypotheses was

15. SUBJECT TERMS

Sequential hypothesis testing, changepoint detection, rapid intrusion detection, automatic threat detection and classification.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Alexander Tartakovsky
UU	UU	UU	UU		19b. TELEPHONE NUMBER 213-740-2450

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

Report Title

Final Report: General Multidecision Theory: Hypothesis Testing and Changepoint Detection with Applications to Homeland Security

ABSTRACT

The objective was to develop general theories of sequential hypothesis testing and quickest change detection for complex multi-population stochastic models, as well as to apply these theories to automatic threat detection and classification with low false alarm and miss-classification rates. More specifically, we addressed complex stochastic models that include: multi-population/multi-channel models; multi-hypothesis scenarios; general stochastic models with non-stationary and dependent observations; prior uncertainty. Within a limited duration of the project (about 8 months) a background for a general theory of testing multiple composite hypotheses was established, and certain particular examples were considered.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

<u>Received</u> <u>Paper</u>

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

<u>Received</u> <u>Paper</u>

09/08/2013 3.00 G. Fellouris, A.G. Tartakovsky. Unstructured Sequential Testing in Sensor Networks, Proceedings of the 52nd IEEE Conference on Decision and Control. 10-DEC-13, . : ,

TOTAL: 1

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

<u>Received</u> <u>Paper</u>

TOTAL:

Number of Manuscripts:

Books

Received		Book
07/29/2013	1.00	A.G. Tartakovsky, I. Nikiforov, M. Basseville. Sequential Analysis: Hypothesis Testing and Change-Point Detection, Boca Raton, Florida, USA: Chapman & Hall/CRC, (12 2013)
07/29/2013	2.00	A.G. Tartakovsky. Rapid Detection of Attacks in Computer Networks by Quickest ChangepointDetection Methods, London: Imperial College Press, (12 2013)

TOTAL: **2**

Received	Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

NAME	PERCENT_SUPPORTED	Discipline
Greg Sokolov	0.10	
FTE Equivalent:	**0.10**	
Total Number:	**1**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
Georgios Fellouris	0.00
FTE Equivalent:	**0.00**
Total Number:	**1**

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Alexander Tartakovsky	0.00	
FTE Equivalent:	**0.00**	
Total Number:	**1**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:...... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):...... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:...... 0.00

Names of Personnel receiving masters degrees

NAME
Total Number:

Names of personnel receiving PHDs

NAME	
Greg Sokolov (awarded May 2014)	
Total Number:	**1**

Names of other research staff

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

See the attached file

Technology Transfer

None

GENERAL MULTIDECISION THEORY:
HYPOTHESIS TESTING AND CHANGEPOINT DETECTION
WITH APPLICATIONS TO HOMELAND SECURITY

FINAL TECHNICAL REPORT
GRANT # W911NF-13-1-0073
REPORT # FINAL-1-14
DATES COVERED: 03/22/2013–11/28/2013

AUTHOR:
Alexander Tartakovsky

PERFORMING ORGANIZATION:

Department of Mathematics
University of Southern California
Los Angeles, CA 90089

Principal Investigator: Alexander Tartakovsky

SPONSORING/MONITORING AGENCY:

The U.S. Army Research Office
Mathematical Sciences Division
P.O. Box 12211
RTP, NC 27709

POC: Dr. Mou-Hsiung Chang

Contents

1. SUMMARY OF ADDRESSED TASKS AND ACCOMPLISHMENTS

We have directed our efforts towards solving the problems associated with Objective 1: *Development of General Multidecision Theory of Sequential Hypothesis Testing* and specifically Task 1: *Development of Nearly Optimal Multidecision Sequential Rules for Testing Multiple Composite Hypotheses.* In particular, we proposed three sequential tests of composite hypotheses and formulated general conditions under which these tests are asymptotically optimal when error probabilities approach zero. Theses conditions are related to the strong law of large numbers (SLLN) for the log-likelihood ratios of hypotheses and rates of convergence in the SLLN. The general theory covers a variety of popular models such as Markov and hidden Markov models and their generalizations.

2. MAIN RESULTS

2.1. General Theory of Sequential Testing Multiple Composite Hypotheses

2.1.1. The Problem and Structure of Sequential Tests

Consider the following general scenario of testing multiple composite hypotheses associated with non-iid stochastic models. Let $(\Omega, \mathscr{F}, \mathscr{F}_n, \mathsf{P}_{\boldsymbol{\theta}})$, $n = 1, 2 \dots$, be a filtered probability space with standard assumptions about monotonicity of the σ-algebras $\mathscr{F}_n$. The vector parameter $\boldsymbol{\theta} = (\theta_1, \dots, \theta_\ell)$ belongs to a subset $\tilde{\Theta}$ of ℓ-dimensional Euclidean space. The sub-σ-algebra $\mathscr{F}_n = \mathscr{F}_n^X = \sigma(\mathbf{X}_1^n)$ of $\mathscr{F}$ is generated by the stochastic process $\mathbf{X}_1^n = (X_1, \dots, X_n)_{n \geqslant 1}$ observed up to time n. The hypotheses to be tested are "$H_i:\ \boldsymbol{\theta} \in \Theta_i$", $i = 0, 1, \dots, N$ ($N \geqslant 1$), where Θ_i are disjoint subsets of $\tilde{\Theta}$. We will also suppose that there is an *indifference* zone $\mathrm{I}_{\mathrm{in}} \in \tilde{\Theta}$ in which there are no constraints on the probabilities of errors imposed. The indifference zone, where any decision is acceptable, is usually introduced keeping in mind that the correct action is not critical and often not even possible when the hypotheses are too close, which is perhaps the case in most, if not all, practical applications. However, in principle I_{in} may be an empty set. The probability measures $\mathsf{P}_{\boldsymbol{\theta}}$ and $\mathsf{P}_{\tilde{\boldsymbol{\theta}}}$ are assumed to be locally mutually absolutely continuous. By $p_{\boldsymbol{\theta}}(X_n|\mathbf{X}_1^{n-1})$, $n \geqslant 1$ we denote corresponding conditional densities which may depend on n.

A multihypothesis sequential test $\delta = (T, d)$ consists of the pair (T, d), where T is a stopping time with respect to the filtration $\{\mathscr{F}_n\}_{n \geqslant 0}$, and $d = d_T(\mathbf{X}_1^T) \in \{0, 1, \dots, N\}$ is an $\mathscr{F}_T$-measurable (terminal) decision rule specifying which hypothesis is to be accepted once observations have stopped (the hypothesis H_i is accepted if $d = i$ and rejected if $d \neq i$, i.e., $\{d = i\} = \{T < \infty,\ \delta \text{ accepts } H_i\}$). The quality of a sequential test is judged on the basis of its error probabilities and expected sample sizes or more generally on the moments of the sample size. Let $\alpha_{ij}(\delta, \boldsymbol{\theta}) = \mathsf{P}_{\boldsymbol{\theta}}(d = j)\mathbb{1}_{\{\boldsymbol{\theta} \in \Theta_i\}}$ ($i \neq j$, $i, j = 0, 1, \dots, N$) be the probability of accepting the hypothesis H_j by the test δ when the true value of the parameter $\boldsymbol{\theta}$ is fixed and belongs to the subset Θ_i and let $\beta_i(\delta, \theta) = \mathsf{P}_{\boldsymbol{\theta}}(d \neq i)\mathbb{1}_{\{\boldsymbol{\theta} \in \Theta_i\}}$ be the probability of rejecting the hypotheses H_i when it is true. Introduce the following two classes of tests

$$
\begin{aligned}
\mathbb{C}(||\alpha_{ij}||) &= \left\{\delta : \sup_{\boldsymbol{\theta} \in \Theta_i} \alpha_{ij}(\delta, \boldsymbol{\theta}) \leqslant \alpha_{ij},\ i, j = 0, 1, \dots, N,\ i \neq j\right\}, \\
\mathbb{C}(\boldsymbol{\beta}) &= \left\{\delta : \sup_{\boldsymbol{\theta} \in \Theta_i} \beta_i(\delta, \boldsymbol{\theta}) \leqslant \beta_i,\ i = 0, 1, \dots, N\right\}
\end{aligned}
\tag{1}
$$

for which maximal error probabilities do not exceed the given numbers α_{ij} and β_i.

The Generalized Matrix Sequential Likelihood Ratio Test. Define the generalized LR statistics

$$\hat{\Lambda}_n^i = \frac{\sup_{\boldsymbol{\theta}\in\Theta} \prod_{k=1}^n p_{\boldsymbol{\theta}}(X_k|\mathbf{X}_1^{k-1})}{\sup_{\boldsymbol{\theta}\in\Theta_i} \prod_{k=1}^n p_{\boldsymbol{\theta}}(X_k|\mathbf{X}_1^{k-1})} = \frac{\prod_{k=1}^n p_{\boldsymbol{\theta}_n^\star}(X_k|\mathbf{X}_1^{k-1})}{\sup_{\boldsymbol{\theta}\in\Theta_i} \prod_{k=1}^n p_{\boldsymbol{\theta}}(X_k|\mathbf{X}_1^{k-1})}, \quad i = 0, 1, \dots, N, \tag{2}$$

where $\boldsymbol{\theta}_n^\star = \arg\sup_{\boldsymbol{\theta}\in\Theta} p_{\boldsymbol{\theta}}(\mathbf{X}_1^n)$ is the MLE estimator. The first multihypothesis test, which we will refer to as the *Multihypothesis Generalized Sequential Likelihood Ratio Test* (MGSLRT), is of the form

$$\text{stop at the first } n \geqslant 1 \text{ such that for some } i \quad \hat{\Lambda}_n^j \geqslant A_{ji} \quad \text{for all } j \neq i \tag{3}$$

and accept the (unique) H_i that satisfies these inequalities, where A_{ij} are positive and finite numbers (thresholds).

The Adaptive Matrix Sequential Likelihood Ratio Test. Let $\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}_n(X_1, \dots, X_n)$ be an estimator of $\boldsymbol{\theta}$ (not necessarily the MLE). If in conditional density $p_{\boldsymbol{\theta}}(X_k|\mathbf{X}_1^{k-1})$ for the k^{th} observation given the previous data $\mathbf{X}_1^{k-1} = (X_1, \dots, X_{k-1})$ we replace the parameter by the estimate $\hat{\boldsymbol{\theta}}_{k-1}$ built upon the sample $\mathbf{X}_1^{k-1}$ that includes $k-1$ observations, then $p_{\hat{\boldsymbol{\theta}}_{k-1}}(X_k|\mathbf{X}_1^{k-1})$ is still a viable probability density, in contrast to the case of the GLR approach where $p_{\hat{\boldsymbol{\theta}}_n}(X_k|\mathbf{X}_1^{k-1})$ is not a probability density anymore for $k \leqslant n$. Therefore, the statistic

$$\check{\Lambda}_n(\boldsymbol{\theta}_i) = \prod_{k=1}^n \frac{p_{\hat{\boldsymbol{\theta}}_{k-1}}(X_k|\mathbf{X}_1^{k-1})}{p_{\boldsymbol{\theta}_i}(X_k|\mathbf{X}_1^{k-1})} = \check{\Lambda}_{n-1}(\boldsymbol{\theta}_i) \times \frac{p_{\hat{\boldsymbol{\theta}}_{n-1}}(X_n|\mathbf{X}_1^{n-1})}{p_{\boldsymbol{\theta}_i}(X_n|\mathbf{X}_1^{n-1})} \tag{4}$$

is a viable likelihood ratio, and it is the nonnegative $\mathsf{P}_{\boldsymbol{\theta}_i}$-martingale with unit expectation, since $\mathsf{E}_{\boldsymbol{\theta}_i}[\check{\Lambda}_n(\boldsymbol{\theta}_i)|\mathbf{X}_1^{n-1}] = \check{\Lambda}_{n-1}^*(\boldsymbol{\theta}_i)$. Therefore, one can use Wald's likelihood ratio identity for finding bounds on error probabilities if $\Lambda_n^*(\boldsymbol{\theta}_i)$ is used instead of the LR with the true parameter value $\boldsymbol{\theta}$. Because of exactly this very convenient property as well as of the simple recursive structure (4) the hypothesis tests based on the adaptive LRs with one-stage delayed estimators represent a very attractive alternative to the GLR tests as well to the mixture-based tests introduced below.

Define the statistics

$$\check{\Lambda}_n^i = \frac{\prod_{k=1}^n p_{\hat{\boldsymbol{\theta}}_{k-1}}(X_k|\mathbf{X}_1^{k-1})}{\sup_{\boldsymbol{\theta}\in\Theta_i} \prod_{k=1}^n p_{\boldsymbol{\theta}}(X_k|\mathbf{X}_1^{k-1})}, \quad i = 0, 1, \dots, N. \tag{5}$$

The second multihypothesis test, which we will refer to as the *Multihypothesis Adaptive Sequential Likelihood Ratio Test* (MASLRT), has the form

$$\text{stop at the first } n \geqslant 1 \text{ such that for some } i \quad \check{\Lambda}_n^j \geqslant A_{ji} \quad \text{for all } j \neq i \tag{6}$$

and accept the (unique) H_i that satisfies these inequalities.

It is convenient to re-write the MASLRT (6) in the following form. Introducing the statistics

$$\ell_n^* = \sum_{k=1}^n \log p_{\hat{\theta}_{k-1}}(X_k|\mathbf{X}_1^{k-1}), \quad \ell_n^i = \sup_{\theta\in\Theta_i} \sum_{k=1}^n \log p_\theta(X_k|\mathbf{X}_1^{k-1})$$

the MASLRT can be written as

$$T^* = \min_{0\leqslant i\leqslant N} T_i^*, \qquad d^* = i \quad \text{if} \quad T^* = T_i^*, \tag{7}$$

where

$$T_i^* = \inf\left\{ n \geqslant 1 : \ell_n^* \geqslant \max_{\substack{0\leqslant j\leqslant N\\ j\neq i}} [\ell_n^j + a_{ji}] \right\}, \quad a_{ij} = \log A_{ij}, \quad i = 0, 1, \dots, N. \tag{8}$$

The Weighted (Mixture-based) Sequential Likelihood Ratio Test. Yet another approach is to use mixtures of LRs (weighted LRs) in test constructions. This approach was proposed by Wald [6] in his seminal work on the SPRT and its extensions to two composite hypotheses.

Define the weighted LRs

$$\bar\Lambda_n^i = \frac{\int_\Theta \prod_{k=1}^n p_{\boldsymbol\theta}(X_k|\mathbf{X}_1^{k-1})\, W(\mathrm{d}\boldsymbol\theta)}{\int_{\Theta_i} \prod_{k=1}^n p_{\boldsymbol\theta}(X_k|\mathbf{X}_1^{k-1}) W_i(\mathrm{d}\boldsymbol\theta)}, \quad i = 0, 1, \dots, N, \tag{9}$$

where the weight functions $W(\boldsymbol\theta)$, $W_i(\boldsymbol\theta)$, $i = 0, 1, \dots, N$, are not necessarily normalized to 1. If the weights are normalized to 1, then they can be regarded as probability distributions. Let $\{A_{ij}\}$ $(i \neq j)$ be positive numbers. The multihypothesis weighted SLRT (MWSLRT) $\bar\delta = (\bar T, \bar d)$ is of the form

$$\text{stop at the first } n \geqslant 1 \text{ such that for some } i \quad \bar\Lambda_n^j \geqslant A_{ji} \quad \text{for all } j \neq i \tag{10}$$

and accept the H_i that satisfies these inequalities.

Taking logarithms and writing

$$\bar\ell_n = \log \int_\Theta \prod_{k=1}^n p_\theta(X_k|\mathbf{X}_1^{k-1})\, W(\mathrm{d}\theta), \quad \bar\ell_n^i = \log \int_{\Theta_i} \prod_{k=1}^n p_\theta(X_k|\mathbf{X}_1^{k-1}) W_i(\mathrm{d}\theta),$$

the MWSPRT can be also expressed as

$$\bar T = \min_{0\leqslant i\leqslant N} \bar T_i, \qquad \bar d = i \quad \text{if} \quad \bar T = \bar T_i \tag{11}$$

where

$$\bar T_i = \inf\left\{ n \geqslant 1 : \bar\ell_n \geqslant \max_{\substack{0\leqslant j\leqslant N\\ j\neq i}} [\bar\ell_n^j + \log A_{ji}] \right\}, \quad i = 0, 1, \dots, N. \tag{12}$$

2.1.2. Probabilities of Errors

One of the most important issues is to obtain upper bounds and approximations for error probabilities of the introduced tests. However, we do not know how to upper-bound the error probabilities of the MGSLRT and the MWSLRT. The reason is that the statistics $\hat\Lambda_n^i$ are not likelihood ratios anymore so that the change-of-measure argument (Wald's likelihood ration identity) cannot be applied. Some asymptotic approximations still can be obtained in the iid case for ℓ-dimensional exponential families using large and moderate deviations:

$$\sup_{\theta\in\Theta_i} \mathsf{P}_\theta(\hat d = j) = \frac{(\log A_{ji})^{\ell/2}}{A_{ji}} + O(1) \quad \text{as} \quad \min_{ij} A_{ij} \to \infty. \tag{13}$$

(cf. Chan and Lai [1]). In the general non-iid case this is still an open problem.

As we mentioned above, in this respect the MASLRT has a big advantage over the MGSLRT to the expense of some loss of performance due to one-stage delayed estimators. Write $\alpha_{ij}^*(\boldsymbol{\theta}) = \mathsf{P}_{\boldsymbol{\theta}}(d^* = j)\mathbb{1}_{\{\boldsymbol{\theta}\in\Theta_i\}}$ for the error probabilities of the MASLRT. We now show that $\sup_{\boldsymbol{\theta}\in\Theta_i} \alpha_{ij}^*(\boldsymbol{\theta}) \leqslant 1/A_{ij}$, $i \neq j$, so that $A_{ij} = 1/\alpha_{ij}$ implies $\delta^* \in \mathbb{C}(||\alpha_{ij}||)$.

Theorem 1. *Let* $\alpha_{ij}^*(\boldsymbol{\theta}) = \mathsf{P}_{\boldsymbol{\theta}}(d^* = j)\mathbb{1}_{\{\boldsymbol{\theta}\in\Theta_i\}}$ *and* $\beta_i^*(\boldsymbol{\theta}) = \mathsf{P}_{\boldsymbol{\theta}}(d^* \neq i)\mathbb{1}_{\{\boldsymbol{\theta}\in\Theta_i\}}$, $i = 0, 1, \dots, N$ *be the error probabilities of the MASLRT* $\delta^* = (d^*, T^*)$. *The following inequalities hold:*

(i) $\sup_{\boldsymbol{\theta}\in\Theta_i} \alpha_{ij}^*(\boldsymbol{\theta}) \leqslant 1/A_{ij}$ *for* $i, j = 0, 1, \dots, N$, $i \neq j$;

(ii) $\sup_{\boldsymbol{\theta}\in\Theta_i} \beta_i^*(\boldsymbol{\theta}) \leqslant \sum_{j\neq i} A_{ij}^{-1}$ *for* $i = 0, 1, \dots, N$.

Proof. Since $\{d^* = j\} = \{T^* = T_j^*\}$ implies $\{T_j^* < \infty\}$, we have

$$\alpha_{ij}^*(\boldsymbol{\theta}) = \mathsf{E}_{\boldsymbol{\theta}}\mathbb{1}_{\{d^*=j\}} \leqslant \mathsf{E}_{\boldsymbol{\theta}}\mathbb{1}_{\{T_j^*<\infty\}} = \mathsf{E}_{\boldsymbol{\theta}}\left[\mathbb{1}_{\{T_j^*<\infty\}}\check{\Lambda}_{T_j^*}(\boldsymbol{\theta})/\check{\Lambda}_{T_j^*}(\boldsymbol{\theta})\right] \quad \text{for all } \boldsymbol{\theta} \in \Theta_i.$$

By the definition of T_j^*, $\check{\Lambda}_{T_j^*}^i \geqslant e^{a_{ij}}$ and clearly $\check{\Lambda}_{T_j^*}(\boldsymbol{\theta}) \geqslant \check{\Lambda}_{T_j^*}^i$ for all $\boldsymbol{\theta} \in \Theta_i$. Therefore, for all $\boldsymbol{\theta} \in \Theta_i$,

$$\alpha_{ij}^*(\boldsymbol{\theta}) \leqslant \mathsf{E}_{\boldsymbol{\theta}}\left[\mathbb{1}_{\{T_j^*<\infty\}}\check{\Lambda}_{T_j^*}(\boldsymbol{\theta})/\check{\Lambda}_{T_j^*}(\boldsymbol{\theta})\right] \leqslant e^{-a_{ij}}\mathsf{E}_{\boldsymbol{\theta}}\left[\mathbb{1}_{\{T_j^*<\infty\}}\check{\Lambda}_{T_j^*}(\boldsymbol{\theta})\right] = e^{-a_{ij}},$$

where the last equality follows from the Wald likelihood ratio identity. This proves (i).

Part (ii) follows immediately from the fact that

$$\beta_i^*(\boldsymbol{\theta}) \leqslant \sum_{j\neq i}\mathsf{P}_{\boldsymbol{\theta}}(T_j^* < \infty) \leqslant \sum_{j\neq i} e^{-a_{ij}}$$

and the proof is complete. □

Therefore, we have the following important implications:

$$A_{ij} = 1/\alpha_{ij} \Longrightarrow \delta^* \in \mathbb{C}(||\alpha_{ij}||); \tag{14}$$

$$A_{ij} = A_i = N/\beta_i \Longrightarrow \delta^* \in \mathbb{C}(\boldsymbol{\beta}). \tag{15}$$

We note that generally there are no such inequalities for the MGSLRT and the MWSLRT.

2.1.3. Near Optimality

The developed asymptotic hypothesis testing theory is based on the SLLN and rates of convergence in the strong law for the LLR processes, specifically by strengthening the strong law into the r-quick version.

Definition 1. For $r > 0$, the random variable ξ_n is said to converge P-r-*quickly* to a constant C if $\mathsf{E}L_\varepsilon^r < \infty$ for all $\varepsilon > 0$, where $L_\varepsilon = \sup\{n : |\xi_n - C| > \varepsilon\}$ ($\sup\varnothing = 0$).

Note that $\mathsf{P}(L_\varepsilon < \infty) = 1$ for all $\varepsilon > 0$ is equivalent to the P-a.s. convergence of ξ_n to C.

Write $\lambda_n(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) = \log \frac{d\mathsf{P}^n_{\boldsymbol{\theta}}}{d\mathsf{P}^n_{\tilde{\boldsymbol{\theta}}}} = \sum_{k=1}^n \log \frac{p_{\boldsymbol{\theta}}(X_k|\mathbf{X}_1^{k-1})}{p_{\tilde{\boldsymbol{\theta}}}(X_k|\mathbf{X}_1^{k-1})}$ for the log-likelihood ratio (LLR) process. Assume that there exist positive and finite numbers $I(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})$ such that

$$\frac{1}{n}\lambda_n(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) \xrightarrow[n\to\infty]{\mathsf{P}_{\boldsymbol{\theta}}-r-\text{quickly}} I(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) \quad \text{for all } \boldsymbol{\theta}, \tilde{\boldsymbol{\theta}} \in \Theta,\ \boldsymbol{\theta} \neq \tilde{\boldsymbol{\theta}}. \tag{16}$$

In addition, we certainly need some conditions on the behavior of the estimate $\hat{\boldsymbol{\theta}}_n$ for large n, which should converge to the true value $\boldsymbol{\theta}$ in a proper way. To this end, we require the following condition on the adaptive LLR process:

$$\frac{1}{n}\log \check{\Lambda}_n(\tilde{\boldsymbol{\theta}}) \xrightarrow[n\to\infty]{\mathsf{P}_{\boldsymbol{\theta}}-r-\text{quickly}} I(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) \quad \text{for all } \boldsymbol{\theta}, \tilde{\boldsymbol{\theta}} \in \Theta,\ \boldsymbol{\theta} \neq \tilde{\boldsymbol{\theta}}, \tag{17}$$

so that the normalized by n LLR tuned to the true parameter value and its adaptive version converge to the same constants. In certain cases, but not always, conditions (16) and (17) imply the following conditions

$$\frac{1}{n}\log \check{\Lambda}_n^i \xrightarrow[n\to\infty]{\mathsf{P}_{\boldsymbol{\theta}}-r-\text{quickly}} I_i(\boldsymbol{\theta}) \quad \text{for all } \boldsymbol{\theta} \in \Theta \setminus \Theta_i,\ i = 0, 1, \dots, N, \tag{18}$$

where $I_i(\boldsymbol{\theta}) = \inf_{\tilde{\boldsymbol{\theta}} \in \Theta_i} I(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})$, the minimal "distance" from $\boldsymbol{\theta}$ to the set Θ_i is assumed to be positive for all i. Let

$$J_i(\boldsymbol{\theta}) = \min_{\substack{0\leqslant j\leqslant N\\ j\neq i}} [I_j(\boldsymbol{\theta})/c_{ji}] \ \text{ for } \ \boldsymbol{\theta} \in \Theta_i, \quad J(\boldsymbol{\theta}) = \max_{0\leqslant i\leqslant N} J_i(\boldsymbol{\theta}) \ \text{ for } \ \boldsymbol{\theta} \in \mathrm{I}_{\mathrm{in}}, \tag{19}$$

and

$$J_i^*(\boldsymbol{\theta}) = \min_{\substack{0\leqslant j\leqslant N\\ j\neq i}} [I_j(\boldsymbol{\theta})/c_j] \text{ for } \boldsymbol{\theta} \in \Theta_i, \ \ J^*(\boldsymbol{\theta}) = \max_{0\leqslant i\leqslant N} \min_{\substack{0\leqslant j\leqslant N\\ j\neq i}} [I_j(\boldsymbol{\theta})/c_j] = \max_{0\leqslant i\leqslant N} J_i^*(\boldsymbol{\theta}) \text{ for } \theta \in \mathrm{I}_{\mathrm{in}}, \tag{20}$$

where $c_{ij} = \lim_{\alpha_{\max}\to 0} |\log \alpha_{ij}|/|\log \alpha_{\max}|$, $\alpha_{\max} = \max_{i,j} \alpha_{ij}$, $c_i = \lim_{\beta_{\max}\to 0} |\log \beta_i|/|\log \beta_{\max}|$, $\beta_{\max} = \max_i \beta_i$.

The following theorem establishes uniform asymptotic optimality of the MASLRT in the general non-iid case with respect to moments of the stopping time distribution. The proof is based on the technique developed by Tartakovsky [5] for multiple simple hypotheses. It is very lengthy and therefore omitted. We just mention that it includes a two-step procedure: first to obtain the asymptotic lower bounds for moments of the stopping time distribution $\inf_{\delta \in \mathbb{C}(||\alpha_{ij}||)(\mathbb{C}(\boldsymbol{\beta}))} \mathsf{E}_{\boldsymbol{\theta}}[T]^m$, $\boldsymbol{\theta} \in \Theta_i$, $m > 0$, $i = 0, 1, \dots, N$, and then to show that these lower bounds are attained for the procedure of interest.

Theorem 2 (Asymptotic Optimality)**.** *Assume that r-quick convergence conditions* (16) *and* (18) *are satisfied.*

(i) *If the thresholds A_{ij} are so selected that $\sup_{\boldsymbol{\theta}\in\Theta_i} \alpha^*_{ij}(\boldsymbol{\theta}) \leqslant \alpha_{ij}$ and $\log A_{ij} \sim \log(1/\alpha_{ij})$, in particular $A_{ij} = 1/\alpha_{ij}$, then for $m \leqslant r$ as $\alpha_{\max} \to 0$*

$$\inf_{\delta\in\mathbb{C}(||\alpha_{ij}||)} \mathsf{E}_{\boldsymbol{\theta}} T^m \sim \mathsf{E}_{\boldsymbol{\theta}}[T^*]^m \sim \begin{cases} [|\log \alpha_{\max}|/J_i(\boldsymbol{\theta})]^m & \textit{for all } \ \boldsymbol{\theta} \in \Theta_i \textit{ and } i = 0, 1, \dots, N \\ [|\log \alpha_{\max}|/J(\boldsymbol{\theta})]^m & \textit{for all } \boldsymbol{\theta} \in \mathrm{I}_{\mathrm{in}}, \end{cases} \tag{21}$$

where the functions $J_i(\boldsymbol{\theta}), J(\boldsymbol{\theta})$ are defined as in (19).
(ii) *If the thresholds $A_{ij} = A_i$ are so selected that $\sup_{\boldsymbol{\theta}\in\Theta_i} \beta_i^*(\boldsymbol{\theta}) \leqslant \beta_i$ and $\log A_i \sim \log(1/\beta_i$, in particular $A_i = N/\beta_i$, then for $m \leqslant r$ as $\beta_{\max} \to 0$*

$$\inf_{\delta\in\mathbb{C}(\boldsymbol{\beta})} \mathsf{E}_{\boldsymbol{\theta}} T^m \sim \mathsf{E}_{\boldsymbol{\theta}}[T^*]^m \sim \begin{cases} [|\log\beta_{\max}|/J_i^*(\boldsymbol{\theta})]^m & \text{for all } \boldsymbol{\theta}\in\Theta_i \text{ and } i=0,1,\dots,N \\ [|\log\beta_{\max}|/J^*(\boldsymbol{\theta})]^m & \text{for all } \boldsymbol{\theta}\in \mathrm{I}_{\mathrm{in}}, \end{cases} \tag{22}$$

where the functions $J_i^*(\boldsymbol{\theta}), J^*(\boldsymbol{\theta})$ are defined as in (20).

Consequently, the MASLRT minimizes asymptotically the moments of the sample size up to order r uniformly in $\boldsymbol{\theta}\in\Theta$ in the classes of tests $\mathbb{C}(||\alpha_{ij}||)$ and $\mathbb{C}(\boldsymbol{\beta})$.

This theorem generalizes previous results of Pavlov [4] and Dragalin and Novikov [2] restricted to iid exponential families, and also provides alternative conditions in iid cases that can be often easily checked.

Remark 1. The assertions of Theorem 2 of course also hold for the MGSLRT and MWSPRT when the r-quick convergence conditions (18) are satisfied for the GLR statistics $\hat{\lambda}_n(\Theta_i) = \hat{\ell}_n - \ell_n^i$ and the mixtures $\bar{\lambda}_n^i$. However, we stress that there are no simple upper bounds for the error probabilities of the MGSLRT and the MWSPRT. Furthermore, while for iid exponential families certain asymptotic approximations for the error probabilities can be obtained based on the boundary-crossing framework and large deviations (see Chan and Lai [1], Lorden [3]), for general non-iid models no such results exist.

Remark 2. The assertions of Theorem 2 remain true if the normalization by n in (18) is replaced with the normalization by $\psi(n)$, where $\psi(t)$ is an increasing function, $\psi(\infty) = \infty$, in which case $[|\log\alpha_{\max}|/J_i(\theta)]^m$ in (21) should be replaced with $\Psi([|\log\alpha_{\max}|/J_i(\theta)]^m)$, where Ψ is inverse to ψ, and similarly in (22).

We now consider two interesting examples.

2.1.4. Testing for the Gaussian Mean with Unknown Variance

Consider the Gaussian example assuming that $X_n \sim \mathcal{N}(\mu, \sigma^2)$, $n = 1, 2, \dots$ are iid normal random variables with unknown mean μ and unknown variance σ^2 and the hypotheses are $H_0 : \mu \leqslant \mu_0, \sigma^2 > 0$ and $H_1 : \mu \geqslant \mu_1, \sigma^2 > 0$, where μ_1, μ_0 are given numbers, $\mu_1 > \mu_0$. The variance σ^2 is a nuisance parameter.

In the following we consider a specific case $H_0 : \mu = 0$ and $H_1 : \mu \geqslant \mu_1$ ($\mu_1 > 0$). This problem is of special interest in certain applications. For example, when detecting targets in noise/clutter the observations have the form $X_n = \mu + V_n$ if there is a target and $X_n = V_n$ if there is no target. The value of μ, $\mu > 0$ characterizes the intensity of the signal from the target; V_n is sensor noise or clutter plus noise. Assuming that $\{V_n\}_{n\geqslant 1}$ is zero-mean white Gaussian noise with unknown variance σ^2, we arrive at this problem. In radar applications, μ usually represents the result of the preprocessing by attenuation and matched filtering of the modulated pulses and, also, is not known. The value of $\mu_1 > 0$ is a prespecified limit or cut-off intensity of the target. In this interpretation the value of $q = \mu/\sigma$ represents an unknown signal-to-noise ratio and $q_1 = \mu_1/\sigma$ is a given cut-off signal-to-noise ratio level. Thus, we are dealing with the two-hypotheses problem ($N = 1$) for the two-dimensional exponential model with the parameter $\boldsymbol{\theta} = (\mu, \sigma^2)$ and parameter space $\Theta = [0, \infty) \times (0, \infty)$. Also, $\Theta_0 = \{0\} \times (0, \infty)$, $\Theta_1 = [\mu_1, \infty) \times (0, \infty)$, and $\mathrm{I}_{\mathrm{in}} = (0, \mu_1) \times (0, \infty)$.

We now show that all the conditions of Theorem 2 are satisfied when $\{\hat{\boldsymbol{\theta}}_n\}$ is a sequence of MLEs, which implies the uniform asymptotic optimality of the 2-ASPRT with $a_i = \log(1/\alpha_i)$, $i = 0, 1$.

Let $\widetilde{\boldsymbol{\theta}} = (\widetilde{\mu}, \widetilde{\sigma}^2)$, where $\widetilde{\mu}$ and $\widetilde{\sigma}$ are arbitrary numbers, $\widetilde{\mu} \geqslant 0$, $\widetilde{\sigma} > 0$. Then the LLR $\lambda_n(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}}) = \ell_n(\boldsymbol{\theta}) - \ell_n(\widetilde{\boldsymbol{\theta}}) = \sum_{k=1}^n \log[p_{\boldsymbol{\theta}}(X_k)/p_{\widetilde{\boldsymbol{\theta}}}(X_k)]$ is given by

$$\begin{aligned}\lambda_n(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}}) = \frac{n}{2} \log\left(\frac{\widetilde{\sigma}^2}{\sigma^2}\right) + \frac{\sigma^2 - \widetilde{\sigma}^2}{2\widetilde{\sigma}^2\sigma^2} \sum_{k=1}^n X_k^2 \\ + \frac{\mu\widetilde{\sigma}^2 - \widetilde{\mu}\sigma^2}{\widetilde{\sigma}^2\sigma^2} \sum_{k=1}^n X_k - \frac{\mu^2\widetilde{\sigma}^2 - \widetilde{\mu}^2\sigma^2}{2\widetilde{\sigma}^2\sigma^2} n.\end{aligned} \tag{23}$$

Using (23), it is not difficult to show that

$$I(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}}) = \frac{1}{2}\left\{[(\mu - \widetilde{\mu})^2 + \sigma^2]/\widetilde{\sigma}^2 + \log(\widetilde{\sigma}^2/\sigma^2) - 1\right\}. \tag{24}$$

The minimum

$$\min_{\widetilde{\sigma}>0} I(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}}) = \frac{1}{2} \log\left[1 + (\mu - \widetilde{\mu})^2/\sigma^2\right]$$

is achieved at the point $\widetilde{\sigma}^* = [\sigma^2 + (\mu - \widetilde{\mu})^2]^{1/2}$ and $I_1(\boldsymbol{\theta}) = \min_{\widetilde{\mu} \geqslant \mu_1} \min_{\widetilde{\sigma}>0} I(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})$ and $I_0(\boldsymbol{\theta}) = \min_{\widetilde{\mu} \in \{0\}} \min_{\widetilde{\sigma}>0} I(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})$ equal

$$\begin{aligned} I_1(q) &= \begin{cases} \frac{1}{2}\log[1 + (q_1 - q)^2] & \text{for } 0 \leqslant q < q_1 \\ 0 & \text{for } q \geqslant q_1 \end{cases}, \\ I_0(q) &= \frac{1}{2}\log(1 + q^2) \quad \text{for } q \geqslant 0, \end{aligned} \tag{25}$$

where $q = \mu/\sigma$ and $q_1 = \mu_1/\sigma$. Thus, as expected, the initial two-dimensional hypothesis testing problem is reduced to the equivalent single-parameter testing problem $H_0 : q = 0$ against $H_1 : q \geqslant q_1$ with the parameter space $Q = [0, \infty)$ and subsets $Q_0 = \{0\}$, $Q_1 = [q_1, \infty)$, $\mathrm{I}_{\mathrm{in}} = (0, q_1)$.

Clearly, $I_0(q) > 0$ for $q \in Q_1 + \mathrm{I}_{\mathrm{in}} = (0, \infty)$ and $I_1(q) > 0$ for $q \in Q_0 + \mathrm{I}_{\mathrm{in}} = [0, q_1)$, and hence, $\min[I_0(q), I_1(q)] > 0$ for $q \in \mathrm{I}_{\mathrm{in}} = (0, q_1)$. Also, it is easily verified that $\inf_{q\in Q} \max[I_0(q), I_1(q)c] > 0$ for any $0 < c < \infty$. In fact, the maximum is attained at the point $q^* \in (0, q_1)$ for which $I_0(q^*) = I_1(q^*)c$, and it is a solution of the equation

$$(1 + q^2)^{1/c} = 1 + (q_1 - q)^2. \tag{26}$$

In particular, $q^* = q_1/2$ and $\inf_{q\in Q} \max[I_0(q), I_1(q)c] = \log(1 + q_1^2/4)$ for $c = 1$. Therefore, the conditions related to the minimal Kullback–Leibler (K–L) distances for the corresponding sets hold, and it remains to deal with convergence of the LLR and associated statistics.

Not surprisingly we choose $\hat{\boldsymbol{\theta}}_n = (\hat{\mu}_n, \hat{\sigma}_n^2)$ as the maximum likelihood estimator,

$$(\hat{\mu}_n, \hat{\sigma}_n^2) = \arg \sup_{\substack{\mu\geqslant 0,\\ \sigma^2>0}} \lambda_n(\mu, \sigma^2, \widetilde{\mu}, \widetilde{\sigma}^2),$$

which is of course a combination of the positive part of the sample mean and sample variance, i.e.,

$$\hat{\mu}_n = \max\{0, \overline{X}_n\}, \quad \hat{\sigma}_n^2 = n^{-1} \sum_{k=1}^{n} (X_k - \hat{\mu}_n)^2,$$

where $\overline{X}_n = n^{-1} \sum_{k=1}^{n} X_k$ is the sample mean.

First, since the LLR $\{\lambda_n(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})\}_{n \geqslant 1}$ given by (23) is a random walk with drift $I(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})$ and $\mathsf{E}_{\boldsymbol{\theta}} |\lambda_1(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})|^r < \infty$ for all positive r, it follows that

$$n^{-1} \lambda_n(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}}) \xrightarrow[n\to\infty]{\mathsf{P}_{\boldsymbol{\theta}}\text{-}r\text{-quickly}} I(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}}) \quad \text{for all } r > 0.$$

Hence, the conditions (16) are satisfied with $I(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})$ equal the K–L numbers given by (24) and $\boldsymbol{\theta} = (\mu, \sigma^2)$.

Write

$$\check{\lambda}_n^i = \log \check{\Lambda}_n^i, \quad \check{\lambda}_n(\tilde{\boldsymbol{\theta}}) = \log \check{\Lambda}_n(\tilde{\boldsymbol{\theta}}).$$

Note that $\check{\lambda}_n^i = \check{\lambda}_n - \lambda_n^i$, $i = 0, 1$. Since $X_1, X_2, \dots$ are iid and $\mathsf{E}_{\boldsymbol{\theta}} |X_1|^r < \infty$ for all $r > 0$, the following r-quick convergence conditions hold as $n \to \infty$ under $\mathsf{P}_{\boldsymbol{\theta}}$:

$$\hat{\mu}_n \to \mu, \quad \hat{\mu}_n^2 \to \mu^2, \quad \hat{\sigma}_n^2 \to \sigma^2, \quad \sigma_{n,0}^2 \to \sigma^2 + \mu^2, \quad \forall \mu \geqslant 0, \sigma^2 > 0,$$
$$\mu_{n,1} \to \begin{cases} \mu & \text{if } \mu \geqslant \mu_1 \\ \mu_1 & \text{if } 0 \leqslant \mu < \mu_1 \end{cases}, \quad \sigma_{n,1}^2 \to \begin{cases} \sigma^2 & \text{if } \mu \geqslant \mu_1 \\ \sigma^2 + (\mu - \mu_1)^2 & \text{if } 0 \leqslant \mu < \mu_1. \end{cases}$$

Using these relations, it can be verified that $\mathsf{P}_{\boldsymbol{\theta}}$-$r$-quickly as $n \to \infty$

$$\begin{aligned}
& n^{-1} \lambda_n^* \to \left(\mu^2 + \sigma^2 - \log \sigma^2 - 1\right)/2, \quad \forall \mu \geqslant 0, \sigma^2 > 0; \\
& n^{-1} \lambda_n^0 \to \left[\mu^2 + \sigma^2 - \log(\mu^2 + \sigma^2) - 1\right]/2, \quad \forall \mu \geqslant 0, \sigma^2 > 0; \\
& n^{-1} \lambda_n^1 \to \begin{cases} \left(\mu^2 + \sigma^2 - \log \sigma^2 - 1\right)/2 & \text{if } \mu \geqslant \mu_1, \sigma^2 > 0 \\ \left\{\mu^2 + \sigma^2 + \log[\sigma^2 + (\mu - \mu_1)^2] - 1\right\}/2 & \text{if } 0 \leqslant \mu < \mu_1, \sigma^2 > 0. \end{cases}
\end{aligned}$$

Combining these formulas yields

$$n^{-1} \check{\lambda}_n^i \xrightarrow[n\to\infty]{\mathsf{P}_{\boldsymbol{\theta}}\text{-}r\text{-quickly}} I_i(\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \Theta \setminus \Theta_i, \ i = 0, 1 \ \text{ for all } r > 0, \tag{27}$$

where $I_1(\boldsymbol{\theta}) = \min_{\widetilde{\mu} \geqslant \mu_1} \min_{\widetilde{\sigma} > 0} I(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}}) \equiv I_1(q)$ and $I_0(\boldsymbol{\theta}) = \min_{\widetilde{\mu} \in \{0\}} \min_{\widetilde{\sigma} > 0} I(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}}) \equiv I_0(q)$ are given by

$$I_1(q) = 1/2 \ \log[1 + (q_1 - q)^2] \quad \text{for } 0 \leqslant q < q_1; \quad I_0(q) = 1/2 \ \log(1 + q^2) \quad \text{for } q > 0,$$

$q = \mu/\sigma$, $q_1 = \mu_1/\sigma$ (cf. (25)).

Therefore, the conditions (18) are satisfied with $I_i(\boldsymbol{\theta}) = I_i(q)$. By Theorem 2, the 2-ASPRT is asymptotically uniformly optimal in the sense of minimizing all the positive moments of the stopping time distribution: for all $r \geqslant 1$ as $\alpha_{\max} \to 0$

$$\inf_{\delta \in \mathbb{C}(\alpha_0, \alpha_1)} \mathsf{E}_{\boldsymbol{\theta}} T^r \sim \mathsf{E}_{\boldsymbol{\theta}} [T^*]^r \sim \begin{cases} \{2 |\log \alpha_1| / \log[1 + (q_1 - q)^2]\}^r & \text{if } 0 \leqslant q \leqslant q^* \\ \{2 |\log \alpha_0| / \log[1 + q^2]\}^r & \text{if } q \geqslant q^*. \end{cases}$$

In addition,

$$\inf_{\delta \in \mathbb{C}(\alpha_0,\alpha_1)} \sup_{\boldsymbol{\theta}\in\Theta} \mathsf{E}_{\boldsymbol{\theta}} T^r \sim \mathsf{E}_{\boldsymbol{\theta}^*}[T^*]^r \sim \{2|\log\alpha_0|/\log[1+(q^*)^2]\}^r,$$

where $q^* \in (0, q_1)$ is the solution of the equation (26).

These asymptotics are also true for the GSLRT and the WSPRT if the thresholds are selected so that the logarithms of the error probabilities of these tests are asymptotic to $\log \alpha_i$.

2.1.5. Testing for a Nonhomogeneous AR Sequence

Let

$$X_n = \theta \cdot S_n + \xi_n, \quad n = 1, 2, \dots,$$

where S_n is a deterministic function and $\{\xi_n\}_{n\geqslant 1}$ is a stable first-order AR Gaussian sequence given by the recursion

$$\xi_n = \gamma\, \xi_{n-1} + \zeta_n, \quad n \geqslant 1,$$

where $\zeta_1, \zeta_2, \dots$ are iid Gaussian variables, $\zeta_k \sim \mathcal{N}(0, \sigma^2)$, and $|\gamma| < 1$. For the sake of concreteness we set $\xi_0 = 0$ and $S_0 = 0$, while all the results are true for arbitrary deterministic or random initial conditions. The hypotheses are $H_0 : \theta \leqslant \theta_0$ and $H_1 : \theta \geqslant \theta_1$, where $\theta_0 < \theta_1$ are given numbers. That is, $\Theta = (-\infty, \infty)$, $\Theta_0 = (-\infty, \theta_0]$, $\Theta_1 = [\theta_1, \infty)$, $\mathrm{I}_{\text{in}} = (\theta_0, \theta_1)$. In this case, the LLR can be written in the form

$$\lambda_n(\theta, \widetilde{\theta}) = \frac{\theta - \widetilde{\theta}}{\sigma^2} \sum_{k=1}^n \widetilde{S}_k\, \widetilde{X}_k - \frac{\theta^2 - \widetilde{\theta}^2}{2\sigma^2} \sum_{k=1}^n \widetilde{S}_k^2,$$

where $\widetilde{X}_k = X_k - \gamma X_{k-1}$, $\widetilde{S}_k = S_k - \gamma S_{k-1}$ ($X_0 = S_0 = 0$). Direct computation shows that

$$\mathsf{E}_\theta[\lambda_n(\theta, \widetilde{\theta})] = \frac{(\theta - \widetilde{\theta})^2}{2\sigma^2} \sum_{k=1}^n \widetilde{S}_k^2.$$

Suppose that

$$\lim_{n\to\infty} n^{-1} \sum_{k=1}^n \widetilde{S}_k^2 = \widetilde{S}^2, \tag{28}$$

where $\widetilde{S}^2$ is a positive and finite number. Then, for all $\theta, \widetilde{\theta} \in (-\infty, \infty)$ ($\theta \neq \widetilde{\theta}$),

$$n^{-1}\lambda_n(\theta, \widetilde{\theta}) \to \frac{(\theta - \widetilde{\theta})^2 \widetilde{S}^2}{2\sigma^2} \quad \mathsf{P}_\theta - r - \text{quickly for all } r > 0. \tag{29}$$

Indeed, under P_θ, the whitened observations $\widetilde{X}_n$ can be written as $\widetilde{X}_n = \theta \widetilde{S}_n + \zeta_n$ and the LLR as

$$\lambda_n(\theta, \widetilde{\theta}) = \frac{\theta - \widetilde{\theta}}{\sigma^2} V_n + \frac{(\theta - \widetilde{\theta})^2}{2\sigma^2} \sum_{k=1}^n \widetilde{S}_k^2,$$

where $V_n = \sum_{k=1}^n \widetilde{S}_k \zeta_k$ is a weighted sum of iid normal random variables. Since $\mathsf{E} V_n = 0$ and, by (28), for large n $\mathsf{E} V_n^2 \sim \widetilde{S}^2 n$, it is obvious that there exists a number $\delta < 1$ such that $\mathsf{P}(|V_n| > \varepsilon n) \leqslant O(\delta^n)$, which yields

$$\sum_{n=1}^\infty n^{r-1} \mathsf{P}(|V_n| > \varepsilon n) < \infty \quad \text{for some } \varepsilon > 0 \text{ and all } r > 0. \tag{30}$$

In other words, V_n/n converges to 0 r-quickly, which implies (29). Hence, the conditions (16) hold with $I(\theta,\widetilde{\theta}) = (\theta-\widetilde{\theta})^2\widetilde{S}^2/2\sigma^2$, and it remains to check the conditions (18), where

$$
\begin{aligned}
I_1(\theta) &= \inf_{\widetilde{\theta}\geqslant\theta_1} I(\theta,\widetilde{\theta}) = \frac{(\theta_1-\theta)^2\widetilde{S}^2}{2\sigma^2} \quad \text{for } \theta<\theta_1,\\
I_0(\theta) &= \inf_{\widetilde{\theta}\leqslant\theta_0} I(\theta,\widetilde{\theta}) = \frac{(\theta-\theta_0)^2\widetilde{S}^2}{2\sigma^2} \quad \text{for } \theta>\theta_0.
\end{aligned}
\tag{31}
$$

As the estimate $\hat{\theta}_n$, we of course use the MLE $\sum_{k=1}^n \widetilde{S}_k\widetilde{X}_k/\sum_{k=1}^n \widetilde{S}_k^2$. Moreover let $\hat{\theta}_{n,1} = \max(\theta_1,\hat{\theta}_n)$ and $\hat{\theta}_{n,0} = \min(\theta_0,\hat{\theta}_n)$. Then the statistics $\check{\lambda}_n^i$ can be written as

$$
\check{\lambda}_n^i = \frac{1}{\sigma^2}\sum_{k=1}^n(\hat{\theta}_{k-1}-\hat{\theta}_{n,i})\widetilde{S}_k\widetilde{X}_k - \frac{1}{2\sigma^2}\sum_{k=1}^n(\hat{\theta}_{k-1}^2-\hat{\theta}_{n,i}^2)\widetilde{S}_k^2, \quad i=0,1.
$$

Using an argument similar to that has led to (30) with a minor generalization, we conclude that, r-quickly under P_θ,

$$
\begin{aligned}
&\hat{\theta}_n \to \theta, \quad \hat{\theta}_{n,1} \to \max(\theta_1,\theta), \quad \hat{\theta}_{n,0} \to \max(\theta_0,\theta), \quad \hat{\theta}_n^2 \to \theta^2,\\
&\hat{\theta}_{n,1}^2 \to \max(\theta_1^2,\theta^2), \quad \hat{\theta}_{n,0}^2 \to \max(\theta_0^2,\theta^2), \quad n^{-1}\sum_{k=1}^n \widetilde{S}_k\widetilde{X}_k \to \theta^2\widetilde{S}^2,
\end{aligned}
$$

which after some manipulations yield

$$
n^{-1}\lambda_n^i \xrightarrow[n\to\infty]{\mathsf{P}_\theta\text{-}r\text{-quickly}} I_i(\theta), \quad \theta\in\Theta\setminus\Theta_i,\ i=0,1 \ \text{ for all } r>0,
$$

where the $I_i(\theta)$'s are given by (31).

Thus, by the Theorem 2, the 2-ASPRT is asymptotically optimal, minimizing all the positive moments of the sample size: for all $r\geqslant 1$ as $\alpha_{\max}\to 0$,

$$
\inf_{\delta\in\mathbb{C}(\alpha_0,\alpha_1)} \mathsf{E}_\theta T^r \sim \mathsf{E}_\theta[T^*]^r \sim \begin{cases} \{2|\log\alpha_1|/[(q_1-q)^2\widetilde{S}^2]\}^r & \text{if } q\leqslant q^*\\ \{2|\log\alpha_0|/[(q-q_0)^2\widetilde{S}^2]\}^r & \text{if } q\geqslant q^*,\end{cases}
$$

where $q=\theta/\sigma$, $q_i=\theta_i/\sigma$, and q^* is a solution of the equation

$$
|\log\alpha_0|/[(q-q_0)^2 = |\log\alpha_1|/(q_1-q)^2.
$$

In particular, if $S_n=1$, then $\widetilde{S}^2=(1-\gamma)^2$.

3. SCIENTIFIC PERSONNEL SUPPORTED BY THIS PROJECT

1. Alexander Tartakovsky, Associate Director, Research Professor, Center for Applied Mathematical Sciences and Department of Mathematics
2. Greg Sokolov, PhD student, Department of Mathematics
3. Georgios Fellouris, Post-doc, Department of Mathematics

4. REQUIRED NUMERICAL DATA RELATED TO THIS GRANT

(1) List of papers, books and book chapters submitted or published under ARO sponsorship during reporting period

(a) Books, book chapters and manuscripts submitted, but not published:

1. A.G. Tartakovsky, Rapid Detection of Attacks in Computer Networks by Quickest Changepoint Detection Methods, *Chapter in the book: Data Analysis for Network Cyber-Security*, Imperial College Press, 2014 (to be published).
2. A.G. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and ChangePoint Detection*, Chapman & Hall/CRC, 2014 (to be published).

(b) Papers published in peer-reviewed journals: None

(c) Papers published in non-peer-reviewed journals or in conference proceedings:

1. T. Banerjee, V.V. Veeravalli, and A. Tartakovsky, Decentralized Data-Efficient Quickest Change Detection, in *Proc. IEEE ISIT*, Istanbul, Turkey, July 2013.
2. G. Fellouris and A.G. Tartakovsky, Unstructured Sequential Testing in Sensor Networks, *Proceedings of the 52nd IEEE Conference on Decision and Control*, Florence, Italy, December 2013.

(d) Papers presented at meetings, but not published in conference proceedings:

1. A.G. Tartakovsky, Rapid Detection of Attacks in Computer Networks by Quickest Changepoint Detection Methods, *Data Analysis for Network Cyber-Security Workshop*, Heilbronn Institute for Mathematical Research, Bristol, UK, March 25-26, 2013 **(Invited Talk)**.
2. A.G. Tartakovsky, Sequential Hypothesis Tests: Historical Overview and Recent Results, *International Workshop in Sequential Methodologies*, Athens, Georgia, July 17–21, 2013 **(Invited Plenary Lecture)**.

(2) Demographic Data for this Reporting Period:

(a) Number of Manuscripts submitted during this reporting period: 4
(b) Number of Peer Reviewed Papers submitted during this reporting period: 4
(c) Number of Non-Peer Reviewed Papers submitted during this reporting period: 0
(d) Number of Presented but not Published Papers submitted during this reporting period: 2

(3) Demographic Data for the life of this agreement:

(a) Number of Scientists Supported by this agreement: 3
(b) Number of Inventions resulting from this agreement: 0
(c) Number of PhD(s) awarded as a result of this agreement: 0
(d) Number of Bachelor Degrees awarded as a result of this agreement: 0
(e) Number of Patents Submitted as a result of this agreement: 0
(f) Number of Patents Awarded as a result of this agreement: 0
(g) Number of Grad Students supported by this agreement: 1
(h) Number of FTE Grad Students supported by this agreement: 0
(i) Number of Post Doctorates supported by this agreement: 1
(j) Number of FTE Post Doctorates supported by this agreement: 1
(k) Number of Faculty supported by this agreement: 1

(l) Number of Other Staff supported by this agreement: 0
(m) Number of Undergrads supported by this agreement: 0
(n) Number of Master Degrees awarded as a result of this agreement: 0

(4) Student Metrics for graduating undergraduates funded by this agreement

(a) Number of undergraduates funded by your agreement during this reporting period: 0
(b) Number of undergraduate funded by your agreement, who graduated during this period: 0
(c) Number of undergraduates funded by your agreement, who graduated during this period with a degree in a science, mathematics, engineering, or technology field: 0
(d) Number of undergraduates funded by your agreement, who graduated during this period and will continue to pursue a graduate or Ph.D degree in a science, mathematics, engineering, or technology field: 0
(e) Number of undergraduates funded by your agreement, who graduated during this period and intend to work for the Defense Department: 0
(f) Number of undergraduates graduating during this period, who achieved at least a 3.5 GPA based on a scale with a maximum of a 4.0 GPA. (Convert GPAs on any other scale to be an equivalent value on a 4.0 scale.): 0
(g) Number of undergraduates working on your agreement, who graduated during this period and were funded by a DoD funded Center of Excellence for Education, Research or Engineering: 0
(h) Number of undergraduates funded by your agreement, who graduated during this period and will receive a scholarship or fellowship for further studies in a science, mathematics, engineering or technology field: 0

(5) Report of inventions

None.

REFERENCES

[1] H. P. Chan and T. L. Lai. Asymptotic approximations for error probabilities of sequential or fixed sample size tests in exponential families. *Annals of Statistics*, 28(6):1638–1669, Dec. 2000.

[2] V. P. Dragalin and A. A. Novikov. Adaptive sequential tests for composite hypotheses. *Surveys in Applied and Industrial Mathematics*, 6(2):387–398, 1999.

[3] G. Lorden. Nearly optimal sequential tests for exponential families. *Unpublished Manuscript*, 1977.

[4] I. V. Pavlov. Sequential procedure of testing composite hypotheses with applications to the Kiefer-Weiss problem. *Theory of Probability and its Applications*, 35(2):280–292, 1990.

[5] A. G. Tartakovsky. Asymptotic optimality of certain multihypothesis sequential tests: Non-i.i.d. case. *Statistical Inference for Stochastic Processes*, 1(3):265–295, Oct. 1998.

[6] A. Wald. *Sequential Analysis*. John Wiley & Sons, Inc, New York, USA, 1947. ISBN 9780486439129.